What is claimed is:

[Claim 1] A neural network comprising:

a first node;

a second node adapted to receive and process signals from said first node;

a first directed edge between said first node and said second node for transmitting signals from said first node to said second node, wherein said first directed edge is characterized by a first weight;

an output node adapted to receive and process signals from said second node;

a second directed edge between said second node and said output node for transmitting signals from said second node to said output node, wherein said second directed edge is characterized by a second weight;

a plurality of additional nodes between said second node and said output node;

a first plurality of directed edges coupling said second node to said plurality of additional nodes;

a second plurality of directed edges coupling said plurality of additional nodes to said output node;

a third plurality of directed edges coupling signals from nodes among said plurality of additional nodes to other nodes among said plurality of additional nodes that are closer to said output node;

wherein, said first weight has a value that is determined by a process of training said neural network that comprises:

estimating a derivative of a summed input to said output node with respect to said first weight by:

multiplying a signal output by said first node by a value of a derivative of a transfer function of said second node that obtains when training data is applied to said neural network to obtain a first factor;

multiplying said first factor by said second weight to compute a first summand;

for each particular node of the plurality of additional nodes between said second node and said output node, computing an additional summand by multiplying together the first factor, a weight characterizing one of the first plurality of directed edges that couples the second node to the particular node, a weight characterizing one of the second plurality of directed edges that couples the particular node to the output node, and a value of a transfer function of the particular node; and

summing the first summand and the additional summands, wherein, in estimating said derivative, paths from said second node to said output node that involve said third plurality of directed edges are not considered.

[Claim 2] The neural network according to claim 1 wherein said first directed edge, said second directed edge, said first plurality of directed edges and said second plurality of directed edges comprise one or more amplifying circuits.

[Claim 3] The neural network according to claim 1 wherein said first directed edge, said second directed edge, said first plurality of directed edges, and said second plurality of directed edges comprise one or more attenuating circuits.

[Claim 4] The neural network according to claim 1 wherein said first node comprises an input of said neural network.

[Claim 5] The neural network according to claim 1 wherein said first node comprises a hidden processing node of said neural network.

[Claim 6] The neural network according to claim 1 wherein:

said plurality of additional nodes include sigmoid transfer functions.

[Claim 7] The neural network according to claim 1 wherein said process of training said neural network comprises:

(a) applying training data to said neural network, whereby said summed input is generated at said output node;

(b) computing a value of a derivative of an objective function that depends on said derivative of said summed input to said output node with respect to said first weight;

(c) processing said derivative of said objective function with an optimization algorithm that uses derivative information; and

(d) repeating (a)-(c) until a stopping condition is satisfied.

[Claim 8] The neural network according to claim 7 wherein in said process of training said neural network, processing said derivative of said objective function comprises:

using a nonlinear optimization algorithm selected from the group consisting of the steepest descent method, the conjugate gradient method, and the Broyden-Fletcher-Goldfarb-Shanno method.

[Claim 9] The neural network according to claim 7 wherein in said process of training said neural network:

(a)-(b) are repeated for a plurality of training data sets, and an average of said derivatives of said objective function over said plurality of training data sets is used in (c).

[Claim 10] The neural network according to claim 7 wherein in said process of training said neural network:

after (d), setting weights that fall below a predetermined threshold to zero.

[Claim 11] The neural network according to claim 10 wherein:

the objective function is a function of a difference an actual output of said neural network that depends on said summed input to said output node and an expected output; and

the objective function is a continuously differentiable function of a measure of near zero weights.

[Claim 12] The neural network according to claim 11 wherein:

the measure of near zero weights takes the form:

$$U = \sum_{i=1}^{K} e^{-\eta w_i^2}$$

where, W_i is a an ith weight

K is a number of weights in the neural network;

η is a scale factor to which weights are compared.

[Claim 13] A method of training a neural network that comprises:

a first node;

a second node adapted to receive and process signals from said first node;

a first directed edge between said first node and said second node for transmitting signals from said first node to said second node, wherein said first directed edge is characterized by a first weight;

an output node adapted to receive and process signals from said second node;

a second directed edge between said second node and said output node for transmitting signals from said second node to said output node, wherein said second directed edge is characterized by a second weight;

a plurality of additional nodes between said second node and said output node;

a first plurality of directed edges coupling said second node to said plurality of additional nodes;

a second plurality of directed edges coupling said plurality of additional nodes to said output node;

a third plurality of directed edges coupling signals from nodes among said plurality of additional nodes to other nodes among said plurality of additional nodes that are closer to said output node;

the method comprising:

estimating a derivative of a summed input to said output node with respect to said first weight by:

multiplying a signal output by said first node by a value of a derivative of a transfer function of said second node that obtains when training data is applied to said neural network to obtain a first factor;

multiplying said first factor by said second weight to compute a first summand;

for each particular node of the plurality of additional nodes between said second node and said output node, computing an additional summand by multiplying together the first factor, a weight characterizing one of the first plurality of directed edges that couples the second node to the particular node, a weight characterizing one of the second plurality of directed edges that couples the particular node to the output node, and a value of a transfer function of the particular node; and

summing the first summand and the additional summands, wherein, in estimating said derivative, paths from said second node to said output node that involve said third plurality of directed edges are not considered.

[Claim 14] The method of training the neural network according to claim 13 wherein comprising:

(a) applying training data to said neural network, whereby said summed input is generated at said output node;

(b) computing a value of a derivative of an objective function that depends on said derivative of said summed input to said output node with respect to said first weight;

(c) processing said derivative of said objective function with an optimization algorithm that uses derivative information; and

(d) repeating (a)-(c) until a stopping condition is satisfied.

[Claim 15] The method of training the neural network according to claim 14 wherein said derivative of said objective function comprises:

using a nonlinear optimization algorithm selected from the group consisting of the steepest descent method, the conjugate gradient method, and the Broyden-Fletcher-Goldfarb-Shanno method.

[Claim 16] The method of training the neural network work according to claim 14 wherein:

(a)-(b) are repeated for a plurality of training data sets, and an average of said derivatives of said objective function over said plurality of training data sets is used in (c).

[Claim 17] The method of training the neural network according to claim 14 wherein:

after (d), setting weights that fall below a predetermined threshold to zero.

[Claim 18] The method of training the neural network according to claim 17 wherein:

the objective function is a function of a difference an actual output of said neural network that depends on said summed input to said output node and an expected output; and
the objective function is a continuously differentiable function of a measure of near zero weights.

[Claim 19] The method of training the neural network according to claim 18 wherein:

the measure of near zero weights takes the form:

$$U = \sum_{i=1}^{K} e^{-\eta W_i^2}$$

where, W_i is a an ith weight

K is a number of weights in the neural network;

η is a scale factor to which weights are compared.